

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      June 22, 2002, 01:52:40 ; Search time 65.99 Seconds
              (without alignments)
              954.238 Million cell updates/sec
```

Title: US-10-112-331-4  
Perfect score: 1877  
Sequence: 1 LRALVFHGLQYAEIIPKSEI.....RRLDFAFIAYNDWNGCEP 364

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```

Searched:      562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :
1:  sp.archaea:*
2:  sp.bacteria:*
3:  sp.fungi:*
4:  sp.human:*
5:  sp.invertebrate:*
6:  sp.mammal:*
7:  sp.mhc:*
8:  sp.organelle:*
9:  sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1532	81.6	364	1	Q9HHB5	Q9HHB5 pyrococcus
2	1494	79.6	364	17	Q58H106	Q58H106 pyrococcus
3	162.5	8.7	529	16	P74630	P74630 synecocyst
4	149.5	8.0	318	17	Q97BM4	Q97BM4 thermoplasma
5	147	7.6	527	16	Q97GF3	Q97GF3 clostridium
6	142.5	7.6	443	17	Q973F0	Q973F0 sulfobolus
7	136.5	7.3	895	17	Q972N0	Q972N0 sulfobolus
8	130.5	7.0	560	17	Q50094	Q50094 pyrococcus
9	128	6.8	923	16	Q96MD4	Q96MD4 bacillus ha
10	124	6.6	1362	17	Q9V294	Q9V294 pyrococcus
11	122.5	6.5	357	17	Q9HL91	Q9HL91 thermoplasma
12	120.5	6.4	324	17	Q30264	Q30264 archaeoglob
13	119	6.3	602	17	Q9V0M7	Q9V0M7 pyrococcus
14	111	5.9	526	16	Q83377	Q83377 trepanema F
15	109.5	5.8	447	17	Q972D2	Q972D2 sulfobolus
16	106	5.6	526	16	Q53278	Q53278 mycobacteri

17	105.5	5.6	902	17	Q97YX0	Q97YX0 subtilobus
18	104	5.5	704	10	Q9ZVN2	Q9ZVN2 arbidopsis
19	103.5	5.5	458	10	Q9MBD0	Q9MBD0 pyrus pyri
20	103.5	5.5	701	2	Q93MG7	Q93MG7 thiodactilu
21	103	5.5	754	13	Q13131	Q13131 oncorhynch
22	102	5.4	754	13	Q13132	Q13132 oncorhynch
23	100.5	5.4	616	16	Q97MA2	Q97MA2 clostridium
24	100.5	5.4	619	16	Q83182	Q83182 treponema
25	99	5.3	4845	11	Q88738	Q88738 mus musculi
26	98	5.2	503	5	Q17585	Q17585 caenorhabdi
27	98	5.2	764	16	Q67347	Q67347 aquifex ae
28	97.5	5.2	402	2	Q68779	Q68779 yersinia pe
29	97.5	5.2	605	16	Q9ZDN2	Q9ZDN2 rickettsia
30	97.5	5.2	627	10	Q9ZOF1	Q9ZOF1 arbidopsi
31	97.5	5.2	684	10	Q9M254	Q9M254 arbidopsi
32	97.5	5.2	888	16	Q98PD2	Q98PD2 mycoplasma
33	97.5	5.2	1085	16	Q25577	Q25577 helicobact
34	97	5.2	510	16	Q66973	Q66973 aquifex ae
35	97	5.2	531	17	Q97BF8	Q97BF8 thermoplas
36	97	5.2	610	2	Q32582	Q32582 escherichia
37	97	5.2	929	5	Q9VR32	Q9VR32 drosophila
38	96.5	5.1	197	13	Q91BE8	Q91BE8 frugu rubr
39	96	5.1	636	2	Q87139	Q87139 vibrio chol
40	96	5.1	636	2	Q34235	Q34235 vibrio chol
41	96	5.1	1150	16	Q99UY8	Q99UY8 staphyloco
42	95.5	5.1	1134	17	Q90Y58	Q90Y58 pyrococci
43	95.5	5.1	1236	5	Q16928	Q16928 caenorhabdi
44	95	5.1	296	2	Q92597	Q92597 streptomyces
45	95	5.1	394	17	Q90YCO	Q90YCO pyrococcus

## ALIGNMENTS

RESULT	1	
ID	09HHB5	PRELIMINARY; PRT; 364 AA.
AC	09HHB5	
DT	01-MAR-2001 (TEMBLrel. 16, Created)	
DT	01-MAR-2001 (TEMBLrel. 16, Last sequence update)	
DT	01-DEC-2001 (TEMBLrel. 19, Last annotation update)	
DE	ALPHA-GALACTOSIDASE.	
GN	GALA.	
OC	Pyrococcus furiosus.	
OC	Archaea, Euryarchaeota; Thermococcales; Pyrococcus	
OX	NCBI_TaxID=2261;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-DSM 3638;	
RA	Verhees C.H.;	
RL	Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.	
DR	EMBL; AF195244; AAC28455.1;	
DR	InterPro; IPR004300; Glyco_hydro.57.	
DR	Pfam; PF03065; Glyco_hydro.57; 1.	
SQ	SEQUENCE 364 AA; 41545 MW; 3E8E0AF5BDCCF2A5 CRC64;	

Query Match	81.6%	Score 1532	DB 1	Length 364
Best Local Similarity	79.1%	Pred. No. 4,9e-117		
Matches 287	Conservative 38	Mismatches 38	Indels 0	Gaps 0

  

QY	1	LRAVAFHGNLQYAEIPKSEIPKVIKAYIPVETLLIKEIPGLNTGTYTLKFLPRDIID	60
Db	1	MRAVAFHGNLQYAEIPKSEIPKVIKAYPPTISELIRPEGLNTGTSLSLSPDLAA	60
QY	61	LKKGIASDLIELTIGTSYTHAIPPLPLSRVEAQVORDREVKELFELSPKGMPLRELAY	120
Db	61	LKEGIESGLIELTIGTSYTHAIPPLPLSRVEAQVORDREVKENITLIVSPSGMPLRELAY	120
QY	121	DPILPILKMDNGEYVFADEGAMLFSAHILNSAKTPKPLTPYLPHILKAQREKRPYISYLLG	180
Db	121	DPILPILKMDNGEYVFADEGAMLFSAHILNSAKTPKPLTPYLPHILKAQREKRPYISYLLG	180



OY 181 LRELKRAIKLVEGVYTLKAVKDIAVAVNTAVMLGRLPLMPKVAWSIEDKO 240  
 DB 181 LRELKRAIKLVEGVYTLKAVKDIAVAVNTAVMLGRLPLMPKVAWSIEDKO 240  
 OY 241 NLLYGTIDIEFGYRDIAGYRMSVBEGLLEIDELNSLCLPSBLKSHSGELIYRTSSMAP 300  
 DB 241 ELLYGTIDIEFGYRDIAGYRMSVBEGLLEIDELNSLCLPSBLKSHSGELIYRTSSMAP 300  
 OY 301 DKSRLRWEDDEGNARLMLSTYMGDELAFLAENSARGMEPLPERRLDAFRAIYNDWGE 360  
 DB 301 DKSRLRWEDDEGNARLMLSTYMGDELAFLAENSARGMEPLPERRLDAFRAIYNDWGE 360  
 OY 361 NGE 363  
 DB 361 NGE 363

RESULT 2

OY 058106 PRELIMINARY: PRT: 364 AA.  
 AC 058106  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 41.8 KDA PROTEIN PH0368.  
 GN PH0368.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 NCBI\_TaxID=53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-073;  
 RX MEDLINE=98344137; PubMed=9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,  
 Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 Funahashi T., Tanaka T., Kudoh Y., Yamazaki N., Guchi A.,  
 Aoki K.-I., Yoshizawa T., Nakamura Y., Kodo F.T., Horikoshi K.,  
 Masuchi Y., Shizuya H., Kikuchi H.;  
 RT Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.;  
 RL DNA Res. 5:55-76(1998).  
 DR EMBL: AF000002; BAA29442.1;  
 DR InterPro: IPR004300; Glyco\_hydro\_57.  
 DR Pfam: PF03065; Glyco\_hydro\_57; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 364 AA: 41755 MW: 7B4B36A4A975BAD CRC64;

Query Match 79.6%; Score 1494; DB 17; Length 364;  
 Best Local Similarity 75.8%; Pred. No. 6.3e-114;  
 Matches 275; Conservative 49; Mismatches 39; Indels 0; Gaps 0;

OY 1 LRALVHGNLQYAEIPKSEIPKVEKAYIPYETLKEEIPFGINTYGTILKEIPKIID 60  
 DB 1 LRALVHGNLQYAEIPKSEIPKVEKAYIPYETLKEEIPFGINTYGTILKEIPKIID 60  
 OY 1 MRALIFGNLQYAEIPKSEIPKVEKAYIPYETLKEEIPFGINTYGTILKEIPKIID 60  
 DB 1 MRALIFGNLQYAEIPKSEIPKVEKAYIPYETLKEEIPFGINTYGTILKEIPKIID 60  
 OY 61 LVKGISASDIEIIGTSTYTHAILPLPLSRVAVOVRDREKEEFELSPKGFMLPELAY 120  
 DB 61 LVKGISASDIEIIGTSTYTHAILPLPLSRVAVOVRDREKEEFELSPKGFMLPELAY 120  
 OY 121 DPITPAIKDNGEYELFADGEAMLFSALNSAIKPIKPLYPHLIKAKREKFRYISYLLG 180  
 DB 121 DPITPAIKDNGEYELFADGEAMLFSALNSAIKPIKPLYPHLIKAKREKFRYISYLLG 180  
 OY 181 LRELKRAIKLVEGVYTLKAVKDIAVAVNTAVMLGRLPLMPKVAWSIEDKO 240  
 DB 181 LRELKRAIKLVEGVYTLKAVKDIAVAVNTAVMLGRLPLMPKVAWSIEDKO 240  
 OY 241 NLLYGTIDIEFGYRDIAGYRMSVBEGLLEIDELNSLCLPSBLKSHSGELIYRTSSMAP 300  
 DB 241 ELLYGTIDIEFGYRDIAGYRMSVBEGLLEIDELNSLCLPSBLKSHSGELIYRTSSMAP 300

OY 301 DKSRLRWEDDEGNARLMLSTYMGDELAFLAENSARGMEPLPERRLDAFRAIYNDWGE 360  
 DB 301 DKSRLRWEDDEGNARLMLSTYMGDELAFLAENSARGMEPLPERRLDAFRAIYNDWGE 360  
 OY 361 NGE 363  
 DB 361 NGE 363

RESULT 3

OY P74630 PRELIMINARY: PRT: 529 AA.  
 AC P74630  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL 62.1 KDA PROTEIN.  
 GN SL0735.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 Miyajima N., Hiroseawa M., Sugijara M., Sasamoto S., Kimura T.,  
 Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 Tabata S.;  
 RT \*Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.;  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL: D90916; BAA18743.1;  
 DR InterPro: IPR003803; DUF200.  
 DR Pfam: PF02651; DUF200; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 529 AA: 62075 MW: 4C1A45048A784E30 CRC64;

Query Match 8.7%; Score 162.5; DB 16; Length 529;  
 Best Local Similarity 21.4%; Pred. No. 7.1e-05;  
 Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;

OY 3 ALVPHGNLQYAEIPKSEIPKVEKAYIPYETLKEEIPFGINTYGTILKEIPKIID 47  
 DB 3 ALVPHGNLQYAEIPKSEIPKVEKAYIPYETLKEEIPFGINTYGTILKEIPKIID 47  
 OY 7 ALVPHGNLQYAEIPKSEIPKVEKAYIPYETLKEEIPFGINTYGTILKEIPKIID 47  
 DB 7 ALVPHGNLQYAEIPKSEIPKVEKAYIPYETLKEEIPFGINTYGTILKEIPKIID 47  
 OY 48 -----GYTLKFLPK----- 56  
 DB 48 -----GYTLKFLPK----- 56  
 OY 67 VSMRLDPLQORYEALSLQELLANEIYRNENHGLQYLAADYAKFAAIRETWERYDG 126  
 DB 67 VSMRLDPLQORYEALSLQELLANEIYRNENHGLQYLAADYAKFAAIRETWERYDG 126  
 OY 57 DIIDLVGKIASDIEIIGTSTYTHAILPLPLSRVAVOVRDREKEEFELSPKGFMLPELAY 114  
 DB 57 DIIDLVGKIASDIEIIGTSTYTHAILPLPLSRVAVOVRDREKEEFELSPKGFMLPELAY 114  
 OY 127 DLYTAIFKOPDSSNNLEITGCGATHGYRPLMKMYPQAWMAIKYACEHEENFGSPGIM 186  
 DB 127 DLYTAIFKOPDSSNNLEITGCGATHGYRPLMKMYPQAWMAIKYACEHEENFGSPGIM 186  
 OY 115 LPELAYDPIITPAIKDNGEYELFADGEAMLFSALNSAIKPIKPLYPHLIKAKREKFRYISYLLG 191  
 DB 115 LPELAYDPIITPAIKDNGEYELFADGEAMLFSALNSAIKPIKPLYPHLIKAKREKFRYISYLLG 191  
 OY 187 LPELAYDPIITPAIKDNGEYELFADGEAMLFSALNSAIKPIKPLYPHLIKAKREKFRYISYLLG 246  
 DB 187 LPELAYDPIITPAIKDNGEYELFADGEAMLFSALNSAIKPIKPLYPHLIKAKREKFRYISYLLG 246  
 OY 152 -----AIKPIKPLYPHLIKAKREKFRYISYLLGRLKRAIKLV-----EGKVT 197  
 DB 152 -----AIKPIKPLYPHLIKAKREKFRYISYLLGRLKRAIKLV-----EGKVT 197  
 OY 247 QQVWSSQVGGPDVYREYFKDLGWEAEYETIKPYIPNCGKRNIGIKYIKITSRDGGLS 306  
 DB 247 QQVWSSQVGGPDVYREYFKDLGWEAEYETIKPYIPNCGKRNIGIKYIKITSRDGGLS 306  
 OY 198 LKAVKDIKAVVAVNTAV-----MLGI-GRPLMPKVAWSIEDKO-----WT 236  
 DB 198 LKAVKDIKAVVAVNTAV-----MLGI-GRPLMPKVAWSIEDKO-----WT 236  
 OY 307 EKAYTD-----PYMAKKEKAASHASNMVNRQOYGHLSGIMGRPLVSPYDAELFGHMY 362  
 DB 307 EKAYTD-----PYMAKKEKAASHASNMVNRQOYGHLSGIMGRPLVSPYDAELFGHMY 362  
 OY 237 E-----DKDNT-----LNYGTIDIEFGYRDIAGYRMSVBEGLLEIDELNSLCLPSBLKSHSGELIYRTSSMAP 260  
 DB 237 E-----DKDNT-----LNYGTIDIEFGYRDIAGYRMSVBEGLLEIDELNSLCLPSBLKSHSGELIYRTSSMAP 260  
 OY 363 EGPFIDYLFKRSWFDODTFEMTHADYLRGNPDQYCRPSQSSMGKGFHEYWLADNTNA 422  
 DB 363 EGPFIDYLFKRSWFDODTFEMTHADYLRGNPDQYCRPSQSSMGKGFHEYWLADNTNA 422  
 OY 261 -----RMSVBEGLLEIDELNSLCLPSBLKSHSGELIYRTSSMAP 299



## FEATURES

Location/Qualifiers  
1. 221284

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-312A14"

1. 1199

/note="assembly\_name:Contig9"

1300. 2874

/note="assembly\_name:Contig10"

2975. 4657

/note="assembly\_name:Contig11"

4758. 6622

/note="assembly\_name:Contig12"

6723. 10907

/note="assembly\_name:Contig13"

11008. 14713

/note="assembly\_name:Contig14"

14814. 20358

/note="assembly\_name:Contig15"

20459. 28691

/note="assembly\_name:Contig16"

28792. 42633

/note="assembly\_name:Contig17"

42734. 70133

/note="assembly\_name:Contig18"

70234. 106104

/note="assembly\_name:Contig19"

106205. 142909

/note="assembly\_name:Contig20"

vector\_side:left"

143010. 179097

/note="assembly\_name:Contig21"

179198. 221284

/note="assembly\_name:Contig22"

BASE COUNT 71097 a 38619 c 39582 g 70672 t 1314 others

ORIGIN

Query Match 1.9%; Score 21; DB 2; Length 221284;

Best Local Similarity 100.0%; Pred. No. 2.8;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 gcttcctccaaaggattctg 341

|||||

Db 15968 GCTTCTCCAAAGGATTCTG 15988

RESULT 14

LOCUS AR076196 52 bp DNA linear PAT 30-AUG-2000

DEFINITION Sequence 1 from patent US 5958751.

ACCESSION AR076196

VERSION AR076196.1 GI:10002942

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 52)

AUTHORS Murphy, D. and Reid, J.

TITLE .alpha.-galactosidase

JOURNAL Patent: US 5958751-A 1 28-SEP-1999;

FEATURES Location/Qualifiers

1. 52

source /organism="unknown"

BASE COUNT 18 a 10 c 11 g 13 t

ORIGIN

Query Match 1.8%; Score 20; DB 6; Length 52;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tggagcgctgcgtcttcac 21  
|||||

Db 33 TGAGAGCGCTGCTTTCAC 52

RESULT 15

TBILTA124/c

LOCUS TBILTA124 1688 bp mRNA linear INV 15-NOV-1991

DEFINITION T.brucei mRNA for variant surface protein ILTat 1.24.

ACCESSION X56767

VERSION X56767.1 GI:10453

KEYWORDS variant surface glycoprotein.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 1688)

REFERENCE Carrington, M.

AUTHORS Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D. and

TITLE Direct Submision

JOURNAL Submitted (20-NOV-1990) M. Carrington, CAMBRIDGE UNIVERSITY,

DEPARTMENT OF BIOCHEMISTRY, TENNIS COURT ROAD, CAMBRIDGE CB2 1QW,

UK

2 (bases 1 to 1688)

REFERENCE Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D. and

AUTHORS Turner, M.

TITLE Variant specific glycoprotein of Trypanosoma brucei consists of two

domains each having an independently conserved pattern of cysteine

residues

JOURNAL J Mol. Biol. 221 (3), 823-835 (1991)

FEATURES

source

misc\_feature

1. 1688

/organism="Trypanosoma brucei"

/strain="ssp. brucei"

/isolate="WING 209"

/db\_xref="taxon:5691"

1. 18

/note="mini exon"

52. 1596

/codon\_start=1

/product="variant surface glycoprotein ILTat 1.24"

/protein\_id="CAA0086.1"

/db\_xref="GI:10454"

/db\_xref="SWISS-PROT:P26329"

/translation="MVYRNLIQSLVLAIVLIVENTHFGYKELMOPCELTAEIR

KTAGVAKMKNVSDNSPSKLTLETKMKLTTPAKPESKEALTLRALALANTDLRAR

DNANGIDRAVRAVASEAAGALFSGIQTLDHATDGTCTCLSGSGSGNGNANASO

GCKPLALPELTLEDSDYNDVLSDKGPKISPTLNAGOGSGEGGLQAAASGAGATP

GVPSGSGSRINILGAIIVASAQOQPTPDLSDFGSTARNQADTLYGRAHSITELI

AQGRPKQCTEVEYETMKLLAOKRAALDSIKFOLAASTGKTSYKEDENIKTEYFGKT

NIEALMNKVEEKYKADPEDEPSKESKISDINTEEOLOAVDYAVATMLAKQARD

IAKLETEIADQGRKSPENECNKITEEPKCESEKICSHMKYKAEKNCORNSTAKSS

GVPTQOTAGADTTAEKCKGKGRKCKSPCKWEGGCKDSITLANKORALSVASAA

FVALLF"

52. 120

sig\_peptide

121. 1593

/product="variant surface glycoprotein ILTat 1.24"

BASE COUNT 594 a 389 c 389 g 316 t

ORIGIN

Query Match 1.8%; Score 20; DB 3; Length 1688;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 336 attcgctccagagctcg 355

|||||

Db 160 ATTCTGCTGCCAGACTCG 141

Search completed: June 21, 2002, 23:04:10

1  
2  
3

4  
5  
6

7  
8  
9



Mon Jun 24 08:37:17 2002

Job time: 7436 sec

us-10-112-331-3.rge









/product="putative transporter"  
/protein\_id="AA628456.1"  
/db\_xref="GI:11066094"  
/translation="MGNIRRSYIVLLIIMAAFLMADNLLPNTQOIKERIGISFTO  
MGLVSIKATSAITIIIMGLADIKORRKLIVIGLGEICFLTAIYQYIWMOLAM  
RLTGIGISIIIPGYSILADMPERKRGKSYIETAFGFTGLFMIIAGMIAWRT  
PRTIAAPRFLIAPFYIIIEPKRKGKRIKALEKYEYLRINLEIKRSDTK  
NIIILFLOGIIIGTVPWGIILWVLSFLQYTRGMDSATFVLLIIGIASVPSGLG  
VDYEPANRGRGRTITGIAIFLGLMGLASIGLILPLPSLSTVHNGIATISILFIOF  
VSYAPNRAIVSOVNPEDRGTFGLFIMLNDVGKATIGLPLGFLIETKSMGMEP  
LAYOTLLIGALFPCALWIMIRKSPEDDKYKEILKRAEELISKPT"  
complement(3457..4722)  
/gene="bgla"  
/complement(3457..4722)  
/gene="bgla"  
/note="family 1 glycosyl hydrolase"  
/codon\_start=1  
/transl\_table=11  
/product="beta-glucosidase"  
/protein\_id="AA628457.1"  
/db\_xref="GI:11066095"  
/translation="MLKPEEFLGTATAAHOIEGDKNMNDWYIIOIKLPYKSG  
ACNHEFYKEDIOLMASLGYNAVRSISMSPLPEENKFNENAFRRIYDILLANN  
ITPLVLIHFTSPIMPMKGGFLRENNKPKYKAELEKYLATPEEPVYV  
MMGYITAYPPPIKSPFAKRYASMLAHALAYELHKGKQVGIKVPMLPATDK  
BRDKAAERADULEWYFLDAIMSGVYGAFAKAYVPOSDPFIINTYASVSHW  
NPLKFFDPAKLADVSERKTOMGWSYPRGIYIALKASKYGPLYITENGATLDDEN  
RIEFLIOHLYQVHKAIEDLDVGRGYFYSVMYEMRGFEPRLGVEDYETFERP  
RKSAYIYGIAKSKIEKDEILEKYGLSL"  
4779..5762  
/gene="galT"  
4779..5762  
/gene="galT"  
/codon\_start=1  
/transl\_table=11  
/product="putative galactose-1-phosphate  
uridylyltransferase"  
/protein\_id="AA628458.1"  
/db\_xref="GI:11066096"  
/translation="MRELRYNPITGOWIMVAVRRKRRPMPITNCPGSEENGYSY  
KVLILPNRPMISFADPSEEFYKRRARIGOSYIETPDHDKLDKDEPRDMYK  
VVELWENTKTLKRNPHAYALIFRNKGEYSISTHROGLYALPIPLKARLTEN  
SRKYKSNKELPRLIKEKESERIIETENDTVFPPFASWPEKHAIYKHHVOY  
LTDANKIEISLADIAIQTGTITLIEROMPYTMMIFQAPFNKYKHYHIEFYPI  
LRDNGKIKYVAAIGMGTEWETFYDAIPEENARKLEACKRVSQRILNKGCY"  
BASE COUNT 1896 a 1123 c 1174 g 1569 t  
ORIGIN

Query Match 54.6% Score 598; DB 1; Length 5762;  
Best Local Similarity 71.7% Pred. No. 8.6e-167;  
Matches 784; Conservative 0; Mismatches 310; Indels 0; Gaps 0;

QY 2 tgaagagcgcgtcttcaagcgaacccagatgacgaatcccaagagcgaatcc 61  
DB 1057 TGAGAGCACTAGTCTTTCATGGAATTTACATATGACGAATACCAAGACGAATTC 1116  
QY 62 caaagtcataagagaagcatatcccaagtcagacacgaatgaagaagaattc 121  
DB 1117 CAAAGTTATGAAAGACATACCTCCCACTATTTCAGACTCTATGAGAGAGAGATAC 1176  
QY 122 ctttgagctaaacataagcgtatccttaagctccctccgaagatattagacc 181  
DB 1177 CTTTGGCCTTAAACATTAACGGGATACATCAAGCTTCTCCCAAAAGATTGATTGCC 1236  
QY 182 tcaataaagagcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 241  
DB 1237 TCATTAAGAGAGGAAATAGAGAGTGGCCCATATGAGATTTCTTGGAACAATATCTACAG 1296  
QY 242 caatactccctccctccctccctccctccctccctccctccctccctccctcc 301  
DB 1297 CAATCTCCCACTCTCTCCCACTCTCTAGAGTTGAGGCAATAAAGAGATAGAGAG 1356

QY 302 ttaagaagagccttcagagcttcccaagagattcggcgccagagcgcctatg 361  
DB 1357 TGAAGAAGAAATATTTGGAAGTCTCCCAAGAGAGATTTCGCTCCCAATTTGGCATATG 1416  
QY 362 accgaataccctccatcatcgaagagcgaagcgaagcgaagcgaagcgaagcga 421  
DB 1417 ACCCAATTAATTCACAGAAATTTGAGGACACACAAATATTAATATCTCTTCGACAGGGC 1476  
QY 422 aggcagctcttctcagctcatcctcaactcggcgatgaagcaatgaacgcctatc 481  
DB 1477 AGGCTATGCTATTTCTTAACCACTTAACCTGCAATTAACCACTTAACCTTATACC 1536  
QY 482 caacaccttaagaagcccaaggaagcgccttaagtagtaactcaactctccttgcc 541  
DB 1537 CTCATCTAATTAAGCCCAAGAGAGAGGGGTTACTGTACTTAATTAATCTTTTGGCC 1596  
QY 542 tcaagagccttaagaagcgaataaagcgtcttcttgaagtagaagtagaagcga 601  
DB 1597 TTAGAGAGCTCAAAAAGGCCATTAATCTGTCTTTAAGAAAGACACACTAGAGAG 1656  
QY 602 tcaagagcctgaagcgcgtacccgcttggtggtggtggtggtggtggtggtg 661  
DB 1657 TCAAGAAATCGAAGCTATCCGGTATGGTTTCCATCAATCAATCAATCAATCAAT 1716  
QY 662 tcggaagcctcctctatgaactcctaagaagtagtgcgagcgtgtagaggaagaca 721  
DB 1717 CGGGAAGATTTCACCTATTGAATCTTAAGAAAGTCCCAAAATGGGTAAGAAAGAG 1776  
QY 722 acattctctatacgcgcacgaatagatgattgctatagcgaatgacatgacgtaca 781  
DB 1777 AATTTTCTTATGGAACAGACATAGACTTCCTCGATACAGAGATATAGCAGGATATA 1836  
QY 782 gaaatgagtgtagagatattagagtgtagagtgtagagtgtagagtgtagagtg 841  
DB 1837 AGATTCAATTTCCAACTATTGGAGATTATTAAGAGCTTAAAGAGATTAGGCTAC 1896  
QY 842 cctcagagctgaagcagagtgtagagtgtagagtgtagagtgtagagtgtagag 901  
DB 1897 CCAGGAAGATTAAGCATACTAGTGAAGAAAGCTTATCTTAAGAACTCAAGTGGCTCGG 1956  
QY 902 ataagagcttgagatag 961  
DB 1957 ACAAAAGTTTGAGATTGAGACAGAGATGAGAGGAATCAAGATTAAACATGCTTACT 2016  
QY 962 acaatataag 1021  
DB 2017 CCTATATGATGAGAGACTGCTGATTTTACGTGAATAAGTATGCTCCGAGGTGGAGAC 2076  
QY 1022 cctccctgaag 1081  
DB 2077 CTCCTCCGGAAGAGAGATTAGCGCATTTAAAGCTATATACCACTTGGAGAGGTGAAA 2136  
QY 1082 atggggaaccttag 1095  
DB 2137 ATGGGAACATTAG 2150

RESULT 3

AP000002/c 257000 bp DNA linear BCT 06-APR-2000  
LOCUS  
DEFINITION Pyrococcus horikoshii OT3 genomic DNA, 287001-544000 nt. position (277).  
ACCESSION AP000002 AB009475 AB009476 AB009477 AB009478 AB009479 AB009480  
AB009481 AB009482 AB009483 AB009484 BA000001  
VERSION AP000002.1 GI:3236129  
KEYWORDS  
SOURCE Pyrococcus horikoshii (strain:OT3) DNA.  
ORGANISM Pyrococcus horikoshii  
Archaea: Euryarchaeota: Thermococci: Thermococcales:  
Thermococcaceae: Pyrococcus.  
REFERENCE  
1 (sites)  
Kawarabayashi, Y., Sawada, M., Horikawa, H., Halkawa, Y., Hino, Y.,  
Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y.,



```
/gene="PH0319"
```





